DOCUMENT RESUME

ED 402 348 TM 025 982

AUTHOR Kochan, Susan; And Others
TITLE You Can't Judge a High School by Achievement Alone: Preliminary Findings from the Construction of a Behavioral Indicator of High School Effectiveness.
PUB DATE Apr 96
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Attendance; Behavioral Science Research; *Behavior Patterns; *Classification; Discipline; Dropouts; Effective Schools Research; *High Schools; *Participation; *School Effectiveness; Test Construction
IDENTIFIERS *Educational Indicators; Student Engagement

ABSTRACT

A composite behavioral indicator of high school effectiveness was developed that measures the degree to which schools strike a balance between the press for academic excellence and the need to keep all students actively engaged in schooling. This proposed "participation" indicator is based on behavioral outcomes with demonstrated relationships to student achievement: attendance, discipline, and dropout. The study, based on a sample of 310 public schools, used the State Department of Education assessment program (in a moderately sized Southern state) and the state education performance indicator program to construct both achievement and participation school effectiveness indexes. School effectiveness was then assessed using the achievement index alone, the behavior index alone, and both together. The application of these indexes resulted in differing effectiveness classifications for the sample schools. The next planned phase of the research will visit outlier schools to gather qualitative information about how well the indexes reflect actual conditions. The newly developed participation indicator has the potential to enhance school effectiveness research. (Contains 6 tables and 22 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 402 348

You Can't Judge A High School by Achievement Alone: Preliminary Findings From the Construction of a Behavioral Indicator of High School Effectiveness

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Susan Kochan
Louisiana Department of Education

Abbas Tashakkori
Charles Teddlie
Louisiana State University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
SUSAN KOCHAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Though school effects researchers have long been criticized for using achievement data as the sole index of school effectiveness, student performance on standardized achievement tests remains the predominant criterion for measuring school effectiveness (Good & Brophy, 1986; Purkey & Smith, 1983). Similarly, researchers in the field of education indicators continue to focus almost exclusively on student achievement data, despite numerous calls for the development of performance measures that reflect a broader array of schooling outcomes (Oakes, 1989; Porter, 1991; Willms, 1992).

The need for alternative indicators of school performance is particularly acute at the secondary level, because high schools have diffuse goals that go beyond academic preparation (Arnn & Mangieri, 1988; Levine & Lezotte, 1990; Teddlie & Stringfield, 1989). Furthermore, critics have speculated that schools that are deemed "effective" strictly on the basis of mean student achievement may not be uniformly successful in serving the learning needs of *all their students*. They also worry that schools that place too high a premium on academic excellence may inadvertently alienate their lower-achieving students and ultimately force them out of school (Wehlage & Rutter, 1986).

Purpose

The purpose of this study is to construct and test a composite behavioral indicator of high school effectiveness that measures the degree to which schools strike a balance between the press for academic excellence and the need to keep *all their students* actively engaged in schooling. The proposed "participation" indicator is based on three behavioral outcomes with demonstrated relationships to student achievement: student attendance, discipline, and dropout (Crone & Franklin, 1992). The intent was is not to *replace* achievement as the primary index of school performance, but to provide an additional perspective on the schooling process. The researchers recommend using the behavioral indicator in tandem with a more traditional achievement-based school effectiveness index (SEI) in assessing and monitoring school performance.

A second purpose was to construct an indicator in such a way that districts and states could readily assess the performance of all their schools, not just a few sites targeted for intensive site-based research. This feature represents a crucial departure from school effects research, where prominent researchers have found valuable performance information on the atttendance and discipline characteristics of effective secondary

TM025987

BEST COPY AVAILABLE

schools, but only *after* they began conducting intensive site-based research (Coleman, Hoffer & Kilgore, 1982; Rutter, Maughan, Mortimore & Ouston, 1979).

A three-phase exploratory study utilizing both quantitative and qualitative methods was therefore undertaken with the aim of aim of answering two research questions. Are high schools consistently or differentially effective in promoting student achievement and student participation in schooling? If some schools *are* differentially effective in promoting student achievement and participation, what contextual differences exist between schools that are categorized as consistently effective, consistently ineffective, or differentially effective?

Methodology.

Sampling Strategy

The study, which is currently underway in a moderate-sized southern state, is based on a sample of 310 public schools selected from a statewide study population of 338 schools whose grade configurations include grades 9-12. Roughly two dozen schools were deleted from the initial sample when it was determined that they (a) were magnet or laboratory schools and thus had selective entrance criteria that would make their comparison with more traditional schools problematic, or (b) changed grade configuration or were not in operation at any time during the time frame of the entire study.

Data Sources

Two data sources were used to construct the achievement and participation SEIs: (a) the State Department of Education (SDE) assessment program, which oversees the administration of norm- and criterion-referenced tests to public schools statewide, and (b) the SDE education performance indicator program, which constructs and reports 10 performance indicators at the school, district, and state levels.

Phase I: Indicator Construction

Hypotheses. Two hypotheses were advanced at the outset of Phase I— one relating to the construction of the participation indicator, the other to the relationships between the two indexes. First, it was hypothesized that a composite participation indicator derived from three component scores (student attendance, suspension, and dropout) would be preferable to a model that incorporated only two components. The researchers also theorized that the achievement and participation indicators measured two related yet distinct dimensions of high school performance, and therefore would be positively but moderately correlated.

Indicator Construction Strategy. Though student achievement and participation can be measured in many ways, the researchers chose to construct the SEIs from data that were routinely collected at the school level statewide and therefore were readily accessible to researchers and education decision makers. This strategy is also consistent with recommendations that indicators be constructed in such a way as to pose a minimal reporting burden for school and district staff (Blank, 1993; Oakes, 1989; OERI, 1988). The development of composite

indicators that draw upon multiple facets of the educational process has been deemed essential in order to reflect the complexity of the learning process (Elliott, Ralph & Turnbull, 1993).

In Phase I, two composite SEIs were constructed for each of the 310 sample schools—one, a traditional achievement-based index, the other an experimental behavioral indicator. A total of three years of data were analyzed for each index so as to minimize the likelihood that school outcomes were attributable to data error and not school effects (Willms, 1992). Annual achievement and participation indices therefore were constructed for SYs 1991-92, 1992-93, and 1993-94, and a mean score was calculated for each indicator, representing average school performance for the three-year period SY 1991-92 to 1993-94.

Student Achievement. The achievement index that was calculated is a composite indicator based on mean student performance on all five components of a criterion-referenced test that is administered statewide in grades 10 and 11 and serves as a high school exit examination. Three of the five CRTcomponents (mathematics, English language arts, and reading) are administered to students in grade 10, while the remaining components (science and written composition) are administered to 11th graders. Inasmuch as more components are administered at the 10th grade level and more 10th graders are tested than 11th graders, student performance in 10th grade was expected to have a disproportionate impact on the overall composite score. However, the researchers did not consider this a problem insofar as between-school comparisons were concerned, because the effect was the same for all schools. Though performance information on 9th and 12th grade students would have been desirable, no such data were available. Norm-referenced data, though equally desirable, were similarly unavailable.

The selection of a composite score that is based on multiple subject areas and grade levels was made in deliberate response to the oft-cited criticism that student performance on a single subject area test or at a single grade level is too narrow a measure of student achievement for judging the effectiveness of an entire school (Good & Brophy, 1986; Purkey & Smith, 1983; Rowan, Bossert & Dwyer, 1983; Witte & Walsh, 1990).

The achievement indicator that ultimately was constructed is modeled after a composite achievement index developed in the 1980s for a statewide school incentive program (Oescher & Brooks, 1991), and closely resembles SEIs used by school effects researchers to categorize the effectiveness of elementary schools (Teddlie & Stringfield, 1993). In constructing the achievement indicator, student-level data for each of the five CRT components were summed to the school level, averaged, and transformed to z scores. The five standardized component scores were then summed, averaged, and standardized again, yielding a single school-level standard score reflecting mean student performance across all five components

Student Participation. A similar procedure was followed in constructing the composite participation indicator. Grade-level data on the number of students in attendance, the number of students suspended out of school, and the number of students who dropped out were summed to the school level and percentages calculated, and standardized, yielding three standard component scores for grades 9-12 combined. Because prior

research has demonstrated a positive relationship between achievement and attendance but an inverse relationship between achievement and suspension and dropout rates (Crone & Franklin, 1992), the suspension and dropout component scores were reversed and renamed "percent of student discipline" and "percent of student retention." The researchers thereby ensured that the relationships between achievement and the three participation components were in the same direction. Once this was accomplished, the three component scores were summed (in various combinations), averaged, and standardized again to produce a single standard school-level composite participation score in various forms.

Findings. As noted previously, four versions on the participation indicator were initially constructed: one, based on attendance, discipline, and retention data; and three based on some combination of two as opposed to three data components (attendance/discipline, attendance/retention, and discipline/retention). Multiple regression was next employed in order to determine how much variation in the four participation indices could be accounted for by factors that research has demonstrated as closely related to the three components. Those predictor variables are: (a) socioeconomic status, as represented by the percent of students receiving free lunch; (b) ethnicity, the percentage of the student body that are members of ethnic minorities; (c) an urbanicity scale adopted from the U.S. Census that categorizes sites along a five-point continuum ranging from metropolitan to rural; and (d) school size, based on the cumulative enrollment for grades 9-12 combined. Three of the four predictor variables (socioeconomic status, ethnicity, and community type) are clearly outside the influence of policy makers, while the fourth (school size) is difficult to control, particularly during lean economic times (Salganik, 1994).

Table 1. Total Variance in the Effectiveness Indices Accounted For: By Index Type and Year

	R^2 (Unadjusted)			
	1991	1992	1993	Mean 1991-1993
Composite Participation Variable				
1. Attendance, Discipline, Dropout	**.2441**	**.2250**	**.4021**	**.3501**
2. Attendance, Dropout	.2441	.2880	.3299	.3292
3. Attendance, Discipline	.2431	.2204	.3913	.3534
4. Discipline, Dropout	.1326	.0509	.3096	.2205
Composite Achievement Variable				
Achievement	.4716	.3796	.4171	.4650

As noted in Table 1, more of the variation in between-schools scores was accounted for, using all three participation component scores, therefore the three-component model was adopted. The next best model was Version 3, which combined attendance and discipline data, with the model based on discipline and dropout data identified as the weakest. It is impossible to tell from these preliminary analyses how much of the variation that was not accounted for is attributable to data error or to school effects. It bears noting, however, that none of the four variations on the participation index matched the achievement index in terms of the amount of variance that can be accounted for by intake variables.

To test the second hypothesis of a moderate correlation between the participation and achievement indices, Pearson Product Moment correlations were calculated between the two achievement and participation indicators for each year and for the three-year period 1991-92 to 1993-94. As noted in Table 2, the analyses revealed a moderate relationship between the two criterion variables during each cross-sectional comparison; the correlations ranged from a low of .60446 in 1993-94 to a high of .61953 in 1991-92; the correlation between the three-year mean scores was higher still, at .646.

Table 2.

Product-Moment Correlations for the Criterion Variables (Participation and Achievement)

Variable	Product Moment Correlation
1. 1991-92	.61953
2. 1992-93	.61199
3. 1993-94	.60446
4. x̄ Participation, Achievement*	.64600

*3-year mean values for SY 1991-92 to 1993-94.

Phase II: Implementation and Comparison of the Effectiveness Classification Schemes

The purpose of Phase II was to develop and compare effectiveness ratings using three methods of classification: the achievement index alone, the behavioral index alone, and the two indicators in tandem. Effectiveness ratings were calculated for each individual school year and for the three-year period 1991-92 to 1993-94. All three classification schemes employed multiple regression in order to identify schools with higher or lower than expected mean achievement and/or participation, first taking into consideration context variables outside the control of educators (Salganik, 1994). The same four variables that were used in the Phase I regression analyses were used in the Phase II regression (student socioeconomic status, percent minority, community type, and school size).

Hypotheses. At the outset of Phase II, the researchers speculated that most—but not all—schools that are effective in promoting achievement would be consistently effective in promoting participation. They also theorized that the composite achievement indicator would show greater stability over time than would the

composite participation indicator.

Sampling Strategy. As mentioned previously, the two composite indicators were deliberately constructed from readily accessible data on all schools in the study population so that district and state decision makers could assess the performance of all schools, using the same criteria. In keeping with that purpose, effectiveness classifications were calculated for all 310 schools in the study sample.

Results. Findings from Phase II did not uphold the researchers prediction that *most* schools would be consistently classified under the two effectiveness classification schemes. In order to test the hypothesis, all schools were classified along a continuum from consistently effective (i.e., the school shows higher than expected mean student achievement and participation) to consistently ineffective (i.e., the school shows lower than expected mean achievement and participation), based on their three-year mean SEIs.

A key consideration in the implementation of the classification scheme was the decision where to place the cutoff between so-called "ineffective," "average," and "effective schools." Two classification models were tested: one using the .674 sd demarcation recommended by Lang (1991) and Scheerens (1992), the other based on a quartile distribution, with the upper and lower quartiles labeled "effective" and "ineffective," respectively, and the middle two quartiles combined into an "average" category.

As noted in Table 3, at .674 sd, 50 of 310 schools (16%) were classified as "effective" for achievement, 217 (70%) were classified as "average" (i.e., they performed roughly as predicted), and 43 schools (14%) were classified as "ineffective" (performing lower than predicted). The distribution of schools across effectiveness

Table 3.
Frequencies of Ineffective, Average and Effective Schools, by Selection Criterion

	.674 sd	25% / 50% /25%
Achievement		
Ineffective	43 (13.9%)	77 (24.8%)
Average	217 (70%)	149 (48.1%)
Effective	50 (16.1%)	84 (27.1%)
Participation		
Ineffective	31 (10%)	72 (23.2%)
Average	236 (76%)	159 (51.3%)
Effective	43 (13.9%)	79 (25.5%)

classifications was somewhat different for the participation indicator. Forty-three of 310 schools (13.9 %) were classified as "effective," 236 (76 %) were rated "average," and 31 (10%) were categorized "ineffective." When the effectiveness distributions for the two indicators are compared, utilization of a .674 sd cutoff and an effectiveness classification based on achievement alone results in a somewhat larger percentage of schools classified as either effective or average.

Utilization of the 25%/50%/25% split results in more schools classified as "effective" for "ineffective," and fewer schools identified as "average." For example, based on this categorization scheme and using the attendance indicator alone, 84 schools (27.1%) were classified effective, 149 (48.1%) were average, and 77 (24.8%) were ineffective. Using the participation indicator alone, 79 (25.5%) were identified as effective, 159 (51%) were identified average, and 72 (23.2%) were ineffective.

Table 4 describes the impact on effectiveness classification when both the achievement and participation indicators are implemented with a categorization cutoff of .674 sd, while Table 5 presents the effectiveness distribution when the categorization is based on a 25%/50%/25% category split. When the most stringent (.674 sd) classification scheme is used, only one in four schools is classified consistently across the two scales. When the 25%/50%/25% split is utilized, the percentage of schools that are consistently classified roughly doubles (i.e., increases from 25.5% to 47.4%).

Table 4.
Percent of Schools Classified as Ineffective, Average and Effective by Selection Criterion (.674 sd)

	Achievement	Participation	Both Indicators
Ineffective	43 (13.9%)	31 (10%)	14 (4.5%)
Average	217 (70%)	236 (76%)	51 (16.5%)
Effective	50 (16.1%)	43 (13.9%)	14 (4.5%)
Effective (Participation)/ Ineffective (Achievement)	-	-	0 (0%)
Ineffective (Participation)/ Effective (Achievement)	-	-	0 (0%)

Table 5.
Percent of Schools Classified Ineffective, Average and Effective, by Selection Criterion (25%/50%/25% Split)

	Achievement	Participation	Both Indicators
Ineffective	77 (24.8%)	72 (23.2%)	33 (10.6%)
Average	149 (48.1%)	159 (51.3%	80 (25.8%)
Effective	84 (27.1%)	79 (25.5%)	34 (11%)
Effective (Participation)/ Ineffective (Achievement)	-	-	7 (2.3%)
Ineffective (Participation)/ Effective (Achievement)	-	-	8 (2.6%)

It is interesting to note that obviously and consistently effective schools are identified using both the quartile split categorization cutoff and the more stringent .674 sd split; many schools that are consistently effective or ineffective fall at the extreme ends of the effectiveness distribution. When the categorization cutoff is set at .674 sd., no schools are differentially effective for one indicator and ineffective for the other; differentially effective/ineffective schools are identified only when the categorization cut-off is shifted closer to the mean by the 25%/50%/25% split. These findings suggest that if a school is effective on at least one dimension, its performance on the other is not so strong or so weak that the school's performance on that second dimension is comparable to that of a consistently effective or ineffective school.

Comparative Stability of the SEIs. As noted previously, the researchers predicted that the composite achievement indicator would have greater stability over time than would the composite participation indicator, and the findings from Phase II confirmed that prediction.

Pearson Product Moment correlations were calculated between the residuals for the two indicators to determine how stable the indices function over time. As noted in Table 6, the annual and three-year achievement residuals were consistently larger than similar correlations calculated between the participation residuals. As expected, the annual participation scores correlated highly with the three-year mean participation score, but not so highly as the annual achievement scores correlated with the mean achievement score.

Table 6.

Product-Moment Correlations for the Criterion Variables

	Product Moment Correlations							
Residuals	1	2	3	4	5	6	7	8
1. 1991 Participation	1.0							
2. 1992 Participation	**.61956**	1.0						
3. 1993 Participation	**.61124**	**.53832**	1.0					
4. 1991 Achievement	.50464	.42162	.43711	1.0				
5. 1992 Achievement	.34361	.32533	.40815	**.71485**	1.0			
6. 1993 Achievement	.30073	.27979	.36296	**.81395**	**.85286**	1.0		
7. × Participation*	**.87696**	**.85172**	**.82677**	.53477	.41868	.36688	1.0	
8. × Achievement*	.41040	.36899	.43477	**.90472**	**.92466**	**.95439**	.47420	1.0

*3-year mean values for SY 1991-92 to 1993-94.

Several explanations can be offered for the achievement indicator's greater stability. First, it is generally recognized that behavior (in this case student engagement in schooling as reflected by attendance, suspension and dropout data) typically changes before cognitive change becomes evident. Looking longitudinally at school performance, it is possible that some schools became more or less effective for participation due to alterations in school policy, climate, etc. that could have influenced student attendance, suspension, and/or dropout rates. Such change could occur before similar change occurred in student achievement; indeed, it might be said that some degree of change in those outcomes may be prerequisite to noticeable change in overall student achievement.

A greater degree of instability was also expected from the participation indicator for purely practical reasons. As previously noted, the achievement index is based on student performance on a standardized test administered to students in grades 10 and 11, where test administration and data collection methods are standardized and closely scrutinized. On the other hand, the component data used to construct the participation indicator are reported by districts to the SDE. Though the Department in question has strived in recent years to operationalize and standardize data definitions, the state has less authority to ensure that the definitions are applied consistently and the data collected uniformly. Some fluctuation in the participation component data may therefore be attributable to inconsistencies in the way that schools and districts collect and report student behavioral data over time.

Finally, as previously noted, the achievement index is based on student performance at two grade levels only (grades 10 and 11); however, the student participation index reflects student engagement in schooling for grades 9-12 combined. In essence, the achievement indicator makes assumptions about school-wide achievement based on the performance of two student cohorts, whereas the participation indicator reflects student

participation through all four high school grade cohorts. When Crone, Lang , Teddlie, and Franklin (1995) experimented with various models for combining CRT component data into composite achievement indices, they found greater consistency among school effectiveness classifications based on components administered to the same cohort of students

Phase III: Qualitative Research at Selected Sites

In Phase I, the research team constructed two composite school effectiveness indices based on achievement and participation, respectively. In Phase II, it was demonstrated that the application of those indices to a sample of schools would result in differing effectiveness classifications for some schools. More intriguing still was the finding that, utilizing a 25%/50%/25% split, some schools would be identified as "effective" on one indicator, but "ineffective" on the other. What the researchers were unable to determine through the conclusion of Phase II, however, was which of the classification methods better depicted actual conditions at the school site: the traditional achievement index alone, the experimental behavioral index, or the two indices in combination.

In Phase III, the team will visit eight outlier cases in order to gather qualitative evidence aimed at (a) determining how accurately the participation and achievement SEIs appear to reflect actual conditions at the school site and (b) describing the climate of schools that have been variously classified as consistently or differentially effective. A variety of qualitative methods will be used to compile case studies on the eight Phase III schools, including interviews with school administrators (principal, assistant principal(s), guidance counselor), a faculty-wide survey, a focus group with faculty members, and a student focus group.

At this point in time, eight outlier cases have been targeted for Phase III site visits : four larger, urban/suburban schools, and four smaller, rural schools. One large urban and one small rural school will be visited during late April/early May 1996 in each of four effectiveness categories (consistently effective, consistently ineffective, high participation/low achievement, and low participation/high achievement)

Importance of the Study

School effects research has long demonstrated that effective secondary schools are characterized by higher than expected attendance and achievement and lower than expected student misbehavior and dropout rates (Coleman, Hoffer & Kilgore, 1983; Rutter et al, 1979). Such findings have emerged in the course of intensive, site-based qualitative research in small samples of schools that had previously been identified as effective, based on achievement data alone. The participation indicator, if validated, can offer researchers an opportunity to weigh the effectiveness of large numbers of schools, using multiple criteria. It therefore has implications for both school effectiveness research and education performance indicator research.

References

Ann, J.W. & Mangieri, J. (1988). Effective leadership in effective schools: A survey of principals' attitudes. *NASSP Bulletin*, 72(505), 1-7.

Blank, R.K. (1993). Developing a system of education indicators: Selecting, implementing, and reporting indicators. *Educational Evaluation and Policy Analysis*, 15(1), 65-80.

Coleman, J.S., Hoffer, T., & Kilgore (1983). Cognitive outcomes in public and private schools. *Sociology of Education, 55*, 65-77.

Crone, L.J. & Franklin, B.J. (1992). Progress Profiles. Paper presented at the annual meeting of the American Educational Research Association.

Crone, L.J., Lang, M.J., Franklin, B.J. & Halbrook, M. (1994). Composite versus component scores: Consistency of school effectiveness classification. *Applied Measurement in Education*, 7(4), 303-321.

Crone, L.J., Lang, M.H., Teddlie, C. & Franklin, B.J. Achievement measures of school effectiveness: Comparison of model stability across years. Applied Measurement in Education, 8(4), 365-377.

Elliott, E.J., Ralph, J.H. & Turnbull, B.J. (1993). Monitoring the nation's educational health. *Phi Delta Kappan*, 74, 520-523.

Good, T.L. & Brophy, J. (1986). School effects. In M. Wittrock (Ed.), *Third handbook of research on teaching* (pp 570-602). New York: Macmillan.

Lang, M.H. (1991). Effective school status: A methodological study of classification consistency. Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Levine, D.U. & Lezotte, L.W. (1990). Context differences: Grade level, socioeconomic status, and rural schools. *Unusually effective schools: A review and analysis of research and practice*. Madison, WI: The National Center for Effective Schools Research and Development.

Oakes, J. (1989). What educational indicators? The case for assessing school context. *Educational Evaluation and Policy Analysis*, 11(2), 181-199.

Office of Educational Research and Improvement (1988). *Creating responsible and responsive accountability systems*. Washington, D.C.: U.S. Department of Education.

Porter, A. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis*, 13(1), 12-29.

Purkey, S.C. & Smith, M.S. (1983). *Effective schools: A review*. The Elementary School Journal, 83(4), 427-452.

Rowan, B., Bossert, S.T., & Dwyer, D.C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, 12(4), 24-31.

Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. with Smith, A. (1979). Fifteen thousand hours: Secondary schools and their effects on children. Cambridge, MA: Harvard University Press.

Salganik, L.H. (1994). Apples and apples: Comparing performance indicators for places with similar demographic characteristics. *Educational Evaluation and Policy Analysis*, 16(2), 125-141.

Scheerens, J. (1992). *Effective schooling: Research, theory, and practice*. London: Cassell.

Teddlie, C. & Stringfield, S. (1989). The ethics of behavior for teachers in four types of school effects. *Ethics of Education*, 9(2), 12-14.

Wehlage, G.G. & Rutter, R.A. (1986). Dropping out: How much do schools contribute to the problem? *Teachers College Record*, 87(3), 374-392.

Willms, J.D. (1992). *Monitoring school performance: A guide for educators*. Washington, D.C.: The Falmer Press.

Witte, J.F. & Walsh, D.J. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis, 12*(2), 188-212.